PHYSICAL REVIEW E **108**, 054119 (2023)

Bound on annealing performance from stochastic thermodynamics, with application to simulated annealing

Yutong Luo,[1,2,*] Yi-Zheng Zhen,[3,4,†] Xiangjing Liu,[2,‡] Daniel Ebler,[5,6,§] and Oscar Dahlsten[7,8,9,‖]
[1]*Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom*
[2]*Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China*
[3]*Hefei National Research Center for Physical Sciences at the Microscale and School of Physical Sciences, University of Science and Technology of China, Hefei 230026, China*
[4]*Shanghai Research Center for Quantum Science and CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China*
[5]*Theory Lab, Central Research Institute, 2012 Labs, Huawei Technology Co. Ltd., Hong Kong Science Park, Hong Kong SAR, China*
[6]*Department of Computer Science, The University of Hong Kong, Pokfulam Road, Hong Kong SAR, China*
[7]*Department of Physics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, China*
[8]*Shenzhen Institute for Quantum Science and Engineering and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China*
[9]*Institute of Nanoscience and Applications, Southern University of Science and Technology, Shenzhen 518055, China*

(Received 20 February 2023; revised 28 July 2023; accepted 13 October 2023; published 13 November 2023)

Annealing is the process of gradually lowering the temperature of a system to guide it towards its lowest energy states. In an accompanying paper [Y. Luo *et al.*, Phys. Rev. E **108**, L052105 (2023)], we derived a general bound on annealing performance by connecting annealing with stochastic thermodynamics tools, including a speed limit on state transformation from entropy production. We here describe the derivation of the general bound in detail. In addition, we analyze the case of simulated annealing with Glauber dynamics in depth. We show how to bound the two case-specific quantities appearing in the bound, namely the activity, a measure of the number of microstate jumps, and the change in relative entropy between the state and the instantaneous thermal state, which is due to temperature variation. We exemplify the arguments by numerical simulations on the Sherrington-Kirkpatrick (SK) model of spin glasses.

DOI: 10.1103/PhysRevE.108.054119

I. INTRODUCTION

Simulated annealing (SA) is a heuristic optimization algorithm to approximate the global minimum of a function in a large search space [1–4]. The algorithm models the physical annealing process in metallurgy [5]: by treating the function to be optimized as the energy landscape of a system, the lowest energy state corresponding to the global minimum can be evolved via a gradually cooling process. SA has been massively used in a wide variety of real applications [6–8] and has stimulated novel development in hardware [9,10].

The ideal SA finds the global optimum of any function from a quasistatic cooling process, requiring infinite annealing time [11]. However, in a realistic setting, the cooling schedule is performed in a finite time. As a result, real SA algorithms do not guarantee the global optimum but only find it probabilistically. Significant efforts have been made to investigate the convergence of real SA algorithms [12–14], to analyze the effect of finite-length cooling schedules [15–18], and to gauge the performance of SA in specific problems [19,20]. In an accompanying paper [21], we derived an analytical bound on the performance of annealing by using recent methods from finite-time stochastic thermodynamics [22–29]. The resulting bound holds for all cooling schedules.

In this paper, we provide the full details of the results in Ref. [21] and show explicitly how the theorems and methods in finite-time stochastic thermodynamics can be adapted to quantify the SA performance. We derive a method for bounding the activity in SA, a measure of the number of microstate jumps. We also show how to bound the change in relative entropy between the state and the instantaneous thermal state which is due to temperature variation. Our arguments involve two conjectures which we describe and justify, and which we hope may spark further research.

The paper is organized as follows. In Sec. II, we introduce key concepts in stochastic thermodynamics that we shall use. In Sec. III, we prove the bound proposed in Ref. [21] which characterizes the performance of any (physical or simulated) annealing process. In Sec. IV, we apply this bound to SA and discuss in detail the assessment of SA performance from accessible parameters. Finally, in Sec. V, we conclude our results.

II. STOCHASTIC THERMODYNAMICS

We start by briefly introducing the stochastic thermodynamical framework of annealing processes, using standard terminology of, e.g., Refs. [23,26,27].

---

[*] yutong.luo21@imperial.ac.uk
[†] zhenyizheng@ustc.edu.cn
[‡] liuxj@mail.bnu.edu.cn
[§] ebler.daniel1@huawei.com
[‖] oscar.dahlsten@cityu.edu.hk

©2023 American Physical Society

There is a system of interest. The system has N energy levels denoted as $\{E_i\}_{i=1}^N$. The statistical state of the system is defined as a vector $p(t) = [p_1(t), \ldots, p_N(t)]$, where $p_i(t)$ is the probability of the system on the energy level E_i at time t. The average energy of the system is

$$E_p := \sum_{i=1}^{N} p_i E_i. \tag{1}$$

In annealing, the system of interest is in contact with a heat bath whose temperature T can be controlled to vary in time. If the system is not in thermal equilibrium, it will then thermalize, that is, the state of the system will evolve towards the thermal Gibbs' state $\gamma = [\gamma_i, \ldots, \gamma_N]$, with $\gamma_i = \exp(-\beta E_i)/Z$, where $Z = \sum_i \exp(-\beta E_i)$ is the partition function and $\beta = 1/T$ is the inverse temperature (The Boltzmann constant k_B is taken to be 1.) In annealing, the thermal state $\gamma(t)$ is time dependent due to the varying temperature $T(t)$.

The dynamics of a thermalization process are described by a master equation [26,30,31]:

$$\dot{p}_i(t) = \sum_{j(\neq i)} \Gamma_{ij}(t)p_j(t) - \Gamma_{ji}(t)p_i(t), \tag{2}$$

where $\Gamma_{ij}(t)$ is the generator satisfying $\sum_i \Gamma_{ij}(t) = 0$, $\forall i$ and $\Gamma_{ij}(t) \geqslant 0$, $\forall i \neq j$. In addition, we suppose that the detailed balance condition is satisfied in the process, i.e.,

$$\Gamma_{ij}(t)\gamma_j(t) = \Gamma_{ji}(t)\gamma_i(t), \ \forall i, j, \tag{3}$$

which is commonly assumed in stochastic thermodynamics (See, e.g., Refs. [22,23,27].)

As the system evolves, there are changes in its energy and entropy. It is common to break the change in (average) energy E_p into two parts:

$$\begin{aligned} dE_p = d\left(\sum_i p_i E_i\right) &= \sum_i E_i dp_i + \sum_i p_i dE_i \\ &:= dQ + dW, \end{aligned} \tag{4}$$

where Q is called the heat added to the system and W is called the work done to the system. In annealing, the energy spectrum is unchanged and $dE_p = dQ$. The thermodynamic entropy of the system (again for $k_B = 1$),

$$S_p := -\sum_i p_i \ln p_i, \tag{5}$$

also changes due to the evolution of p. Note that we shall make no distinctions between the thermodynamic entropy and Shannon entropy since the two quantities are the same within our setting. For the thermal state γ transitioning to a new thermal state at an infinitesimally different temperature, as in quasistatic annealing, one can show $\beta\, dQ = dS_p$. If the state is not a thermal state, one has $dS_p \geqslant \beta\, dQ$. The quantity

$$\dot{\Sigma}(t) = \dot{S}_p(t) - \beta(t)\dot{Q}(t) \tag{6}$$

is thus non-negative in the case where the system is not in thermal equilibrium. $\dot{\Sigma}$ is commonly termed the *entropy production* rate. The entropy production when two systems 1 and 2 (like the system and the bath) interact is by definition $dS_1 + dS_2$, in line with early ideas of entropy being akin to a fluid existing in and flowing between different systems and possibly being created [32]. Here the bath is in a thermal state, such that $dS_1 + dS_2 = dS_1 + \beta(t)dQ_2 = dS_1 - \beta(t)dQ_1$, justifying the name *entropy production* rate for $\dot{\Sigma}$ (see, e.g., Ref. [27] for more).

To characterize the performance of an annealing protocol run in finite time τ, we consider the 1-norm distance [22–24] to the final thermal state $\gamma(\tau)$:

$$L_t := \sum_i |p_i(t) - \gamma_i(\tau)|. \tag{7}$$

The probability of not having a globally optimal state at time τ when $T(\tau) = 0$, p(nonoptimal) is bounded by L_τ. To show this, notice that the 1-norm distance is twice the total variation distance [31] such that

$$\begin{aligned} \tfrac{1}{2}L_\tau &= \max_{\text{events}} |p(\text{events}) - \gamma(\text{events})| \\ &\geqslant |p(\text{nonoptimal}) - \gamma(\text{nonoptimal})| \\ &= p(\text{nonoptimal}), \end{aligned} \tag{8}$$

where in the last line we used the fact that $\gamma(\text{nonoptimal}) = 0$ at zero temperature. L_τ thus can be considered as the performance error of the annealing. In the next section, an upper bound of L_τ in terms of L_0 for a general annealing process will be given.

III. GENERAL BOUND ON ANNEALING IN FINITE TIME

We then consider such an annealing process, where a system initially in thermal equilibrium with a heat bath and the bath temperature is continuously decreased from $T_\mathrm{i} = T(0)$ to $T_\mathrm{f} = T(\tau)$ within a finite time τ. In the accompanying paper [21], we derived a general limit on the annealing performance error represented by L_τ. Here, we provide the details of the derivation in the following three subsections.

A. Splitting the relative entropy

A critical challenge in evaluating the effectiveness of an annealing process is to measure the deviation of the system state, $p(\tau)$, from the reference thermal state $\gamma(\tau)$. While the 1-norm distance, L_τ, is a suitable metric for this purpose, its direct calculation can be difficult. However, relative entropy, a comparison tool widely adopted in information theory to compare two probability distributions [33], proves to be a good alternative to start with, due to its deep connection to nonequilibrium thermodynamics. The relative entropy between probability distributions $p(t)$ and $\gamma(t)$ is defined as

$$S(p(t)||\gamma(t)) = \sum_i p_i(t) \ln \frac{p_i(t)}{\gamma_i(t)}. \tag{9}$$

Its close relationship with free energy is apparent via the expression

$$F(t) - F_{\text{eq}}(t) = T(t)S(p(t)||\gamma(t)), \tag{10}$$

where the nonequilibrium free energy $F(t) := E_p(t) - T(t)S_p(t)$ and the equilibrium free energy $F_{\mathrm{eq}}(t) := E_\gamma(t) - T(t)S_\gamma(t)$ with $E_\gamma(t)$ and $S_\gamma(t)$ being the energy and entropy of the thermal state $\gamma(t)$, respectively.

Taking the time derivative of $S(p||\gamma)$, one gets

$$\begin{aligned} \frac{d}{dt}S(p||\gamma) &= \dot{p}\frac{\partial}{\partial p}S(p||\gamma) + \dot{\gamma}\frac{\partial}{\partial \gamma}S(p||\gamma) \\ &= \dot{p}\frac{\partial}{\partial p}S(p||\gamma) + \dot{\beta}\frac{\partial}{\partial \beta}S(p||\gamma) \\ &= -\dot{\Sigma} + \dot{\mathcal{I}}, \end{aligned} \tag{11}$$

We immediately see that a new rate, $\dot{\mathcal{I}} := \dot{\beta}\frac{\partial}{\partial\beta}S(p||\gamma)$, compared to a fixed temperature scenario, appears at the end of Eq. (11). $\mathcal{I}$ is an annealing-related quantity, arising from the temperature variation. A direct calculation shows that

$$\dot{\mathcal{I}} = \dot{\gamma}\frac{\partial}{\partial\gamma}S(p||\gamma) = -\sum_i \dot{\gamma}_i \frac{p_i}{\gamma_i} = (E_p - E_\gamma)\dot{\beta}, \tag{12}$$

where the expression

$$\dot{\gamma}_i = \dot{\beta}(E_\gamma - E_i)\gamma_i \tag{13}$$

is used in the last equality.

We now comment on the physical meaning of these terms from Eq. (11). From Eq. (6), the entropy production rate, $\dot{\Sigma}$, is nonzero when the system thermalizes towards a fixed thermal state. We imagine that, in ideal annealing processes, the change of the thermal states can be regarded as the switching of the system to heat baths at different temperatures. The disconnection and reconnection of the system with baths should not alter the system state instantaneously, resulting in no entropy production during the change of the reference thermal state. However, the variation of the reference state does affect the value of the relative entropy. This contribution is represented by $\dot{\mathcal{I}}$.

We proceed to split the relative entropy. Integrating both sides of Eq. (11) from 0 to τ gives us

$$S(p(\tau)||\gamma(\tau)) - S(p(0)||\gamma(0)) = -\Sigma(\tau) + \mathcal{I}(\tau), \tag{14}$$

where $\Sigma(\tau) = \int_0^\tau \dot{\Sigma}\, dt$ and

$$\mathcal{I}(\tau) = \int_0^\tau (E_p - E_\gamma)\dot{\beta}\, dt. \tag{15}$$

Since the system is assumed to be in equilibrium initially, which is the most common case in annealing, we have $S(p(0)||\gamma(0)) = 0$ and therefore obtain,

$$S(p(\tau)||\gamma(\tau)) = -\Sigma(\tau) + \mathcal{I}(\tau). \tag{16}$$

Using Pinsker's inequality [34], $L_\tau^2 \leqslant 2S(p(\tau)||\gamma(\tau))$, we arrive at an upper bound on L_τ^2,

$$L_\tau^2 \leqslant 2[-\Sigma(\tau) + \mathcal{I}(\tau)]. \tag{17}$$

B. Bounding entropy production during the annealing

In this subsection, we, inspired by Ref. [22], derive a lower bound for the entropy production $\Sigma(\tau)$ in the varying temperature case. For the clarity of equations, we omit the time dependence of $p_i(t)$ and $\Gamma_{ij}(t)$.

First, we rewrite $\dot{\Sigma}(t)$ as

$$\begin{aligned} \dot{\Sigma}(t) &= -\sum_i \dot{p}_i[\ln p_i + \beta(t)E_i] \\ &= \frac{1}{2}\sum_{i\neq j}(\Gamma_{ij}p_j - \Gamma_{ji}p_i)\ln\frac{\Gamma_{ij}p_j}{\Gamma_{ji}p_i} \\ &\geqslant \sum_{i\neq j}\frac{(\Gamma_{ij}p_j - \Gamma_{ji}p_i)^2}{\Gamma_{ij}p_j + \Gamma_{ji}p_i}, \end{aligned} \tag{18}$$

where we have used Eq. (2), the relation $\beta(t)E_i = -\ln\gamma_i - \ln Z$ in the second line, and the inequality $(x-y)\ln(x/y) \geqslant 2(x-y)^2/(x+y)$ in the third line, as suggested by Ref. [22].

Meanwhile, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \sum_i |\dot{p}_i| &= \sum_{i\neq j}|\Gamma_{ij}p_j - \Gamma_{ji}p_i| \\ &\leqslant \sqrt{\left(\sum_{i\neq j}\frac{(\Gamma_{ij}p_j - \Gamma_{ji}p_i)^2}{\Gamma_{ij}p_j + \Gamma_{ji}p_i}\right)\left(\sum_{i\neq j}\Gamma_{ij}p_j + \Gamma_{ji}p_i\right)} \\ &\leqslant \sqrt{2\dot{\Sigma}(t)\mathcal{A}(t)}, \end{aligned} \tag{19}$$

where we have defined the activity

$$\mathcal{A}(t) := \sum_i \sum_{j(\neq i)} \Gamma_{ij}(t)p_j(t), \tag{20}$$

which is the expected jumping rate among different states at time t. We denote the expected number of jumps during the time interval $[0, \tau]$ as

$$\langle N_{\mathrm{jumps}}\rangle = \int_0^\tau \mathcal{A}(t)dt = \langle\mathcal{A}\rangle_\tau \tau, \tag{21}$$

where $\langle\mathcal{A}\rangle_\tau := \frac{1}{\tau}\int_0^\tau \mathcal{A}(t)dt$. [See also the discussion around Eq. (34) for further explanation of why we call this the expected number of jumps.]

Then, the 1-norm distance between $p(0)$ and $p(\tau)$ satisfies

$$\begin{aligned} \sum_i |p_i(0) - p_i(\tau)| &\leqslant \sum_i \int_0^\tau dt|\dot{p}_i(t)| \\ &\leqslant \int_0^\tau dt\sqrt{2\dot{\Sigma}(t)\mathcal{A}(t)} \\ &\leqslant \sqrt{2\Sigma(\tau)\langle\mathcal{A}\rangle_\tau \tau}. \end{aligned} \tag{22}$$

Equivalently, we obtain

$$\Sigma(\tau) \geqslant \frac{\left(\sum_i |p_i(0) - p_i(\tau)|\right)^2}{2\langle\mathcal{A}\rangle_\tau \tau}. \tag{23}$$

That lower bound on $\Sigma(\tau)$ is termed the *speed limit* from entropy production on the transformation from the initial state $p(0)$ to the final $p(\tau)$.

As a concrete example, consider a quasistatic process, which has zero entropy production [$\Sigma(\tau) = 0$]. Then the speed limit implies that for a finite activity $\langle\mathcal{A}\rangle_\tau$, a transition between two distinct statistical states requires $\tau \to \infty$.

Since $\sum_i |p_i(0) - p_i(\tau)|$ is a distance, by triangular inequality we have

$$\begin{aligned}\sum_i |p_i(0) - p_i(\tau)| \\ &\geqslant \left|\sum_i |p_i(0) - \gamma_i(\tau)| - \sum_i |p_i(\tau) - \gamma_i(\tau)|\right| \\ &= |L_0 - L_\tau|. \end{aligned} \tag{24}$$

Substituting Eq. (24) into Eq. (23), we finally obtain a lower bound of the entropy production for a general annealing process.

$$\Sigma(\tau) \geqslant \frac{(L_0 - L_\tau)^2}{2\langle\mathcal{A}\rangle_\tau \tau}. \tag{25}$$

C. General bound

Combining Eqs. (17) and (25), we have

$$L_\tau^2 \leqslant -\frac{(L_0 - L_\tau)^2}{\langle\mathcal{A}\rangle_\tau \tau} + 2\mathcal{I}(\tau). \tag{26}$$

Equation (26) can be regarded as a unary quadratic inequality for L_τ. Solving this inequality for L_τ gives us

$$L_\tau \leqslant \frac{L_0 + \sqrt{\langle\mathcal{A}\rangle_\tau \tau\left[-L_0^2 + 2\mathcal{I}(\tau)(\langle\mathcal{A}\rangle_\tau \tau + 1)\right]}}{\langle\mathcal{A}\rangle_\tau \tau + 1}. \tag{27}$$

This is the general bound of an annealing process as obtained in Ref. [21]. It is worth noticing that Eq. (27) is applicable for any system energy landscape and any cooling schedule as long as the thermalization dynamics respect the detailed balance condition in Eq. (3). Since the terms inside the square root in Eq. (27) are non-negative given Eq. (26), $\langle\mathcal{A}\rangle_\tau$ and $\mathcal{I}(\tau)$ obey a tradeoff relation:

$$\mathcal{I}(\tau) \geqslant \frac{L_0^2}{2(\langle\mathcal{A}\rangle_\tau \tau + 1)}. \tag{28}$$

This can be understood as a "speed limit" related to $\mathcal{I}(\tau)$, which can only be small if the expected number of jumps $\langle N_{\text{jumps}}\rangle = \langle\mathcal{A}\rangle_\tau \tau$ is large, given that L_0 is fixed.

In the quasistatic limit associated with the protocol duration $\tau \to \infty$, the tightness of our bound [Eq. (27)] for the error L_τ is guaranteed. This can be seen by investigating how our bound on SA performance scales with τ. Let the rate of varying temperature $\dot{T}$ decrease with τ. In this case, $\mathcal{I}(\tau)$ will diminish as τ increases. This can be seen from bounding the expression of $\mathcal{I}(\tau)$ in Eq. (15):

$$\mathcal{I}(\tau) = \int_0^\tau (E_p - E_\gamma)\dot{\beta}\, dt \leqslant \max_{t\in[0,\tau]} (E_p - E_\gamma)\Delta\beta, \tag{29}$$

with $\Delta\beta \equiv \beta_{\text{f}} - \beta_{\text{i}}$ being the maximal range of the inverse temperature. As τ increases, the thermal state of the system changes slower, leading to a smaller $\max_{t\in[0,\tau]}(E_p - E_\gamma)$ that measures the departure from the equilibrium energy. We can define the downward scaling of $\mathcal{I}(\tau)$ for large τ in terms of the big O notation for an asymptotic upper bound: $\mathcal{I}(\tau) = \mathcal{O}(\tau^{-\alpha})$, where $0 < \alpha \leqslant 1$ is some scalar. An upper bound on α is in fact imposed by the tradeoff relation [Eq. (28)] [assuming $\langle\mathcal{A}\rangle_\tau$ is finite (bounded)]. Substituting that upper bound of $\mathcal{I}(\tau)$ into Eq. (27), the upper bound on L_τ, we obtain $L_\tau = \mathcal{O}(\tau^{-\alpha/2})$. This scaling implies the tightness of the bound for quasistatic protocols: when $\tau \to \infty$, the protocol approaches being quasistatic and both L_τ and its bound vanish.

We now further analyze how to evaluate Eq. (27) when using it to bound the performance of simulated annealing (SA), in particular, how to evaluate $\langle\mathcal{A}\rangle_\tau$ and $\mathcal{I}(\tau)$ without calculating the evolution of the statistical state of the system, $p(t)$.

IV. APPLICATION TO SIMULATED ANNEALING

In this section we apply the bound on annealing performance, Eq. (27), to the case of simulated annealing (SA).

SA is an optimization algorithm inspired by real annealing processes. The algorithm treats the cost function as the energy of a system and uses a control parameter called "temperature" to anneal it. The physical principle is that a system at thermal equilibrium has a higher probability of staying in its ground state at lower temperatures. The algorithm seeks to find the minimal cost by gradually lowering the temperature and moving closer to the final thermal equilibrium. Therefore, the success of the optimization depends on how far the system is from this equilibrium state.

In SA, the annealing time τ used in the previous section is replaced by the discrete time steps κ. $L_\kappa \equiv \sum_i |p_i(\kappa) - \gamma_i(\kappa)|$ becomes the error from the target state. As we will show, a discrete-time analog of Eq. (27) provides an upper bound to the error when the SA is run in κ steps.

To illustrate how to evaluate $\langle\mathcal{A}^{\text{D}}\rangle_\kappa$ and $\mathcal{I}(\kappa)$, the discrete version of $\langle\mathcal{A}\rangle_\tau$ and $\mathcal{I}(\tau)$, without referring to the intermediate statistical state of the system, we first briefly set up the framework of the simulated annealing in Sec. IV A and bound $\langle\mathcal{A}^{\text{D}}\rangle_\kappa$ and $\mathcal{I}(\kappa)$ in Sec. IV B and Sec. IV C, respectively. A performance bound of simulated annealing is derived in Sec. IV D. In Sec. IV E we show how $\mathcal{I}(\kappa)$ can be calculated analytically for a one-dimensional (1D) Ising chain. A numerical verification of the tightness of the bound on a Sherrington-Kirkpatrick (SK) model, a fully connected spin glass with Gaussian couplings [1,35], is given in Sec. IV F. Finally, in Sec. IV G, we provide further analytical and numerical discussion on the relaxation behavior of SA by using SK models.

A. Simulated annealing algorithm

In SA, the system state is updated iteratively until it approximately reaches equilibrium. The evolution of the state can be viewed as a discrete-time Markov chain [31]. Suppose that temperature decreases from T_{i} to T_{f} in κ steps, i.e., $T_{\text{i}} = T_0, T_1, \ldots, T_k, \ldots, T_\kappa = T_{\text{f}}$. At step k, let the probability of the system staying in state i be $p_i(k)$. For each iteration step k, the state changes are described by

$$p_i(k+1) = \sum_j P_{ij}(k+1)p_j(k), \tag{30}$$

where $P_{ij}(k+1) = \text{Prob}\{X(k+1) = i|X(k) = j\}$ is the transition matrix with $X(k)$ denoting the system state at step k, satisfying $\sum_i P_{ij}(k) = 1$, $\forall j$ and $P_{ij}(k) \geqslant 0$, $\forall i, j$.

In typical SA algorithms, $P_{ij}(k)$ has the following form:

$$P_{ij}(k) = \begin{cases} \frac{1}{|\mathcal{N}(j)|}A_{ij}(k), & i \in \mathcal{N}(j), \\ 1 - \sum_{l(\neq j)} P_{lj}(k), & i = j, \\ 0 & \text{otherwise}, \end{cases} \tag{31}$$

where $|\mathcal{N}(j)|$ is the size of $\mathcal{N}(j)$, the neighbourhood of the state j in which all states can be reached in one step from j, and the acceptance rate $A_{ij}(k)$ is the probability of accepting such a transition from j to i.

In line with SA simulating a thermalisation process, the detailed balance condition [Eq. (3)] is imposed. Here, we more specifically adopt the acceptance rate from the *Glauber dynamics* which is widely used in SA and spin glass studies [31,36–38],

$$A_{ij}(k) = \frac{1}{1+\exp[\beta(k)\Delta E_{ij}]} = \frac{\gamma_i(k)}{\gamma_i(k)+\gamma_j(k)}, \tag{32}$$

where the energy difference $\Delta E_{ij} = E_i - E_j$.

B. Bounding the iteration-averaged activity $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa$ for SA

We now proceed to bound the activity that appears in Eq. (27) for SA with Glauber dynamics. We first define the discrete-time activity $\mathcal{A}^{\mathrm{D}}$. We then derive an expression for $\mathcal{A}^{\mathrm{D}}$ under Glauber dynamics SA. We proceed to conjecture an upper bound on $\mathcal{A}^{\mathrm{D}}$ based on analytical and numerical evidence.

Since the time evolution in SA [Eq. (30)] is discrete, a discrete version of the activity is needed. We shall use a superscript D to denote *discrete*, and write the discrete-time activity as $\mathcal{A}^{\mathrm{D}}$. In the Appendix, we show that any discrete-time Markov chain in Eq. (30) can always be simulated by a continuous-time master equation in Eq. (2). It is, therefore, natural to define the discrete-time activity as

$$\mathcal{A}^{\mathrm{D}}(k) = \sum_i \sum_{j(\neq i)} P_{ij}(k+1)p_j(k), \tag{33}$$

which leads to the *iteration-averaged* activity $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa := \frac{1}{\kappa}\sum_{k=0}^{\kappa-1} \mathcal{A}^{\mathrm{D}}(k)$. $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa$ equals the $\langle \mathcal{A} \rangle_\tau$ of the corresponding continuous-time Markov chain (see the Appendix) and can be interpreted in terms of the expected number of jumps between micro-states. The right-hand side (RHS) of Eq. (33) is a sum over the probabilities of jumping from state i to $j \neq i$, so we may write $\mathcal{A}^{\mathrm{D}} = p(\text{jump})$. The expected number of jumps over the whole process, by inspection, is $\langle N^{\mathrm{D}}_{\text{jumps}} \rangle = \sum_{k=0}^{\kappa-1} p(\text{jump})$. Thus

$$\langle N^{\mathrm{D}}_{\text{jumps}} \rangle = \langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa \kappa. \tag{34}$$

Eq. (34) is the discrete-time analog of Eq. (21). For the case of continuous time, it is natural to define $\langle N_{\text{jumps}} \rangle = \int_{t=0}^{\tau} \rho(\text{jump}, t)dt$ where $\rho(\text{jump}, t)$, which can be termed the rate of jumping, is the probability density for changing microstates in the infinitesimal time interval dt.

Substituting the general transition matrix $P_{ij}(k)$ of Eq. (31) into Eq. (33), we have

$$\mathcal{A}^{\mathrm{D}}(k) = \sum_i \sum_{j \in \mathcal{N}(i)} \frac{1}{|\mathcal{N}(j)|} A_{ij}(k+1)p_j(k). \tag{35}$$

For simplicity, we denote $|\mathcal{N}(j)| = n_j$ and ignore the step dependence of quantities, writing $q(k) \equiv q$, where q is some quantity. We use prime $'$ to denote the quantity at the next step: $q(k+1) \equiv q'$.

Note that the acceptance rate A'_{ij} defined in Eq. (32) has the property $A'_{ij} + A'_{ji} = 1$. We can therefore rewrite Eq. (35) as

$$\begin{aligned} \mathcal{A}^{\mathrm{D}} &= \sum_i \sum_{j \in \mathcal{N}(i)} \frac{1}{n_j} A'_{ij} p_j \\ &= \frac{1}{2} \sum_i \sum_{j \in \mathcal{N}(i)} \left(A'_{ij}\frac{p_j}{n_j} + A'_{ji}\frac{p_i}{n_i} \right) \\ &= \frac{1}{2} \sum_i \sum_{j \in \mathcal{N}(i)} \frac{p_i}{n_i} + \frac{1}{2} \sum_i \sum_{j \in \mathcal{N}(i)} A'_{ij}\left(\frac{p_j}{n_j} - \frac{p_i}{n_i} \right) \\ &= \frac{1}{2} - \frac{1}{4} \sum_i \sum_{j \in \mathcal{N}(i)} (A'_{ij} - A'_{ji}) \left(\frac{p_i}{n_i} - \frac{p_j}{n_j} \right) \\ &= \frac{1}{2} - \frac{1}{4} \sum_i \sum_{j \in \mathcal{N}(i)} \frac{\gamma'_i - \gamma'_j}{\gamma'_i + \gamma'_j} \left(\frac{p_i}{n_i} - \frac{p_j}{n_j} \right). \end{aligned} \tag{36}$$

Therefore, the activity, interpreted as the jumping probability, deviates from 1/2, the unbiased case, by a number relevant to the order of occupation probabilities for neighboring states. More precisely, we can consider the activity at the thermal state by setting $p_i = \gamma'_i, \ \forall i$, which gives us

$$\begin{aligned} \mathcal{A}^{\mathrm{D}}_\gamma &\equiv \sum_i \sum_{j \in \mathcal{N}(i)} \frac{1}{n_j} A'_{ij}\gamma'_j \\ &= \frac{1}{2} - \frac{1}{4} \sum_i \sum_{j \in \mathcal{N}(i)} \frac{\gamma'_i - \gamma'_j}{\gamma'_i + \gamma'_j} \left(\frac{\gamma'_i}{n_i} - \frac{\gamma'_j}{n_j} \right). \end{aligned} \tag{37}$$

The thermal state, at infinite temperature, becomes uniform (unbiased), such that $\mathcal{A}^{\mathrm{D}}_\gamma = 1/2$, while at zero temperature $\mathcal{A}^{\mathrm{D}}_\gamma = 0$, meaning that any possible jumps are frozen. The analysis of the two extreme cases is intuitive and simple. However, the value of $\mathcal{A}^{\mathrm{D}}_\gamma$ at the intermediate temperatures highly depends on the neighbourhood structure of the energy landscape, making the determination of $\mathcal{A}^{\mathrm{D}}_\gamma$ quite complicated.

One common way to simplify the problem is by considering equal-sized neighborhoods, i.e., $n_i = n, \ \forall i$, as in the Edwards-Anderson model [39] we shall introduce later. Now, by Eq. (37), we have

$$\mathcal{A}^{\mathrm{D}}_\gamma = \frac{1}{2} - \frac{1}{4n} \sum_i \sum_{j \in \mathcal{N}(i)} \frac{(\gamma'_i - \gamma'_j)^2}{\gamma'_i + \gamma'_j} \leqslant 1/2. \tag{38}$$

We conclude that, at least for fixed neighborhood size, $0 \leqslant \mathcal{A}^{\mathrm{D}}_\gamma \leqslant 1/2$ during the annealing process, where equality on both sides holds at zero and infinite temperatures, respectively.

There is significant evidence that $\mathcal{A}^{\mathrm{D}}(k)$ also has a nontrivial upper bound in SA. In SA, the system, initially in a thermal state at some nonzero temperature, keeps evolving toward a cooler thermal state. It is thus reasonable to conjecture that

$$\mathcal{A}^{\mathrm{D}}(k) \leqslant 1/2, \quad 0 \leqslant k \leqslant \kappa. \tag{39}$$

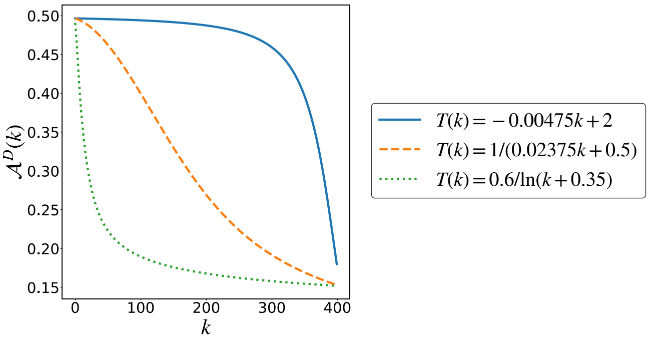

FIG. 1. The activity $\mathcal{A}^{\mathrm{D}}(k)$ in the SA of a 7-spin SK model. Simulations take 400 steps with temperature changing from $T_{\mathrm{i}} = 2$ to $T_{\mathrm{f}} = 0.1$, according to the linear (blue solid line), inversely linear (orange dashed line) and inversely logarithmic (green dotted line) cooling schedules, respectively.

Numerical evidence from simulations of a 7-spin SK model (the formal definition of the SK model will be given in Sec. IV F), is provided in Fig. 1. One sees Eq. (39) is respected. Moreover, there is a decreasing trend of $\mathcal{A}^{\mathrm{D}}(k)$ during the annealing, which coincides with previous numerical results [15] and could be a general feature of SA. As a consequence of the conjecture [Eq. (39)], we have $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa \leqslant 1/2$.

A further, intuitive, route to understanding conjecture [Eq. (39)] makes use of the tendency for lower energy states to have a higher probability in thermal scenarios. For equal-sized neighborhoods, the activity $\mathcal{A}^{\mathrm{D}}$ is given by

$$\mathcal{A}^{\mathrm{D}} = \frac{1}{2} - \frac{1}{4n} \sum_i \sum_{j \in \mathcal{N}(i)} \frac{(\gamma_i' - \gamma_j')(p_i - p_j)}{\gamma_i' + \gamma_j'}. \tag{40}$$

Adopting the concept of *passive states* where the lower energy state has the higher probability of occupation [40–44], we could assume the *passivity in neighborhoods* and define the neighborhood-passivity $\mathcal{P}_i$ for state i,

$$\begin{aligned} \mathcal{P}_i &:= \sum_{j \in \mathcal{N}(i)} \frac{(\gamma_i' - \gamma_j')(p_i - p_j)}{\gamma_i' + \gamma_j'} \\ &= \sum_{j \in \mathcal{N}(i)} \tanh\left[\frac{\beta'}{2}(E_j - E_i)\right](p_i - p_j). \end{aligned} \tag{41}$$

Note that if the two neighboring states p_i and p_j respect passivity, $\tanh[\beta'(E_j - E_i)/2](p_i - p_j) \geqslant 0$. The passivity in neighborhoods thus requires $\mathcal{P}_i \geqslant 0$, $\forall i$. Once this assumption holds during the SA, by Eq. (40) we have $A^{\mathrm{D}} \leqslant 1/2$ for every step k. Consequently, $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa \leqslant 1/2$. This assumption can be justified by noticing that all thermal states are passive. Thus, as the system initially staying in a thermal state evolves towards a new thermal state in SA, we expect the passivity at least in the neighborhoods to be preserved. As is shown in Fig. 2, $\mathcal{P}_i \geqslant 0$, $\forall i$, for three different kinds of cooling schedules, which verifies our passivity assumption and gives new numerical evidence for the conjecture of $\mathcal{A}^{\mathrm{D}}(k) \leqslant 1/2$, and hence $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa \leqslant 1/2$.

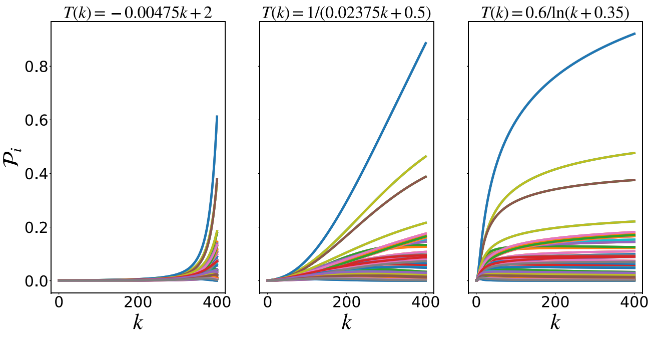

FIG. 2. Verification of the neighborhood-passivity assumption in a 7-spin SK model. The simulation takes 400 steps with temperature changing from $T_{\mathrm{i}} = 2$ to $T_{\mathrm{f}} = 0.1$, according to the linear (left), inversely linear (middle) and inversely logarithmic (right) cooling schedules, respectively. Each curve represents the neighborhood passivity of a state in the state space. All curves are above $\mathcal{P}_i = 0$.

C. Bounding the relative entropy change due to temperature variation $\mathcal{I}(\kappa)$

In this subsection, we will bound the relative entropy change from temperature variation, $\mathcal{I}$. In SA, $\mathcal{I}$ is evaluated by a discrete sum:

$$\mathcal{I}(\kappa) = \sum_{k=1}^{\kappa} [E_p(k) - E_\gamma(k)][\beta(k) - \beta(k-1)], \tag{42}$$

which is shown to be equal to the $\mathcal{I}(\tau)$ of the underlying continuous-time process that simulates the discrete-time evolution in the Appendix. Since $E_\gamma(k)$ is the energy expectation of the thermal state γ, although it apparently acquires the k dependence from a cooling schedule $T(k)$, it is intrinsically a fixed function of temperature for a specific energy landscape. In this sense, we regard it as a problem-specific but dynamics-free quantity. In contrast, the average energy $E_p(k)$ requires information on the instantaneous system state $p(k)$. To access it, one has to solve the dynamics in Eq. (30), which is generally difficult and is the main obstacle to analyzing the performance of SA. In the following, we will try to tackle it by providing an upper bound to $E_p(k)$ in terms of $E_\gamma(k)$. This is done via the use of a partial swap (PS) model, a simple thermalization protocol where the system state is swapped to the thermal state with a probability at each step [28,45,46]. We shall use the superscript PS to denote quantities in the partial swap model. By showing that $E_p(k) \leqslant E_p^{\mathrm{PS}}(k)$ for all k, we will derive a bound of $\mathcal{I}(\kappa)$ as $\mathcal{I}(\kappa) \leqslant \mathcal{I}^{\mathrm{PS}}(\kappa)$.

The evolution of the state of the system p^{PS} in a partial swap model is given by

$$p_i'^{\mathrm{PS}} = p_i^{\mathrm{PS}} + \mu^{\mathrm{PS}}\left(\gamma_i'^{\mathrm{PS}} - p_i^{\mathrm{PS}}\right), \tag{43}$$

where μ^{PS} is the so-called partial swap rate and the time dependence has been neglected and the prime $'$ denotes the next time step. Multiplying E_i on both sides and summing over i gives

$$E_p'^{\mathrm{PS}} = E_p^{\mathrm{PS}} + \mu^{\mathrm{PS}}\left(E_\gamma'^{\mathrm{PS}} - E_p^{\mathrm{PS}}\right), \tag{44}$$

and therefore

$$\mu^{\mathrm{PS}} = \frac{E_p'^{\mathrm{PS}} - E_p^{\mathrm{PS}}}{E_\gamma'^{\mathrm{PS}} - E_p^{\mathrm{PS}}}. \tag{45}$$

Similarly, we can define a *relaxation rate* μ for the actual process in SA as

$$\mu := \frac{E'_p - E_p}{E'_\gamma - E_p}. \tag{46}$$

This relaxation rate has been employed in analyzing the minimal dissipation and designing the optimal cooling schedule in SA [47–49]. Note that, by writing in this form, we require a positive μ. Since in SA, $E_p \geqslant E_\gamma \geqslant E'_\gamma$, we need the assumption that $E_p \geqslant E'_p$. Such descending energy expectation has been witnessed in most uses of SA [15,50,51]. μ^{PS} and μ quantify the relaxation speeds of PS and the actual processes in terms of the energy differences, respectively. Intuitively, $\mu \geqslant \mu^{\mathrm{PS}}$ implies that the corresponding PS process is *slower* than the actual one in SA. This can be formulated by the following proposition.

Proposition 1. Consider that two identical systems initially in the same thermal state are annealed according to the same cooling schedule. Their relaxation dynamics are given by Eq. (30) and Eq. (43), respectively. If $\mu \geqslant \mu^{\mathrm{PS}}$ during the annealing, $E_p \leqslant E_p^{\mathrm{PS}}$ holds all the time.

Proof. Since the two processes correspond to the same energy landscape and cooling schedule, we have $E'^{\mathrm{PS}}_\gamma = E'_\gamma$. Let $\mu \geqslant \mu^{\mathrm{PS}}$, namely,

$$\frac{E'_p - E_p}{E'_\gamma - E_p} \geqslant \frac{E'^{\mathrm{PS}}_p - E^{\mathrm{PS}}_p}{E'_\gamma - E^{\mathrm{PS}}_p}. \tag{47}$$

As a partial swap rate, $\mu^{\mathrm{PS}} \leqslant 1$. The insufficient thermalization leads to $E'^{\mathrm{PS}}_p \geqslant E'_\gamma$, and therefore, μ^{PS} is a monotonically increasing function of E^{PS}_p. If $E^{\mathrm{PS}}_p \geqslant E_p$,

$$\frac{E'^{\mathrm{PS}}_p - E^{\mathrm{PS}}_p}{E'_\gamma - E^{\mathrm{PS}}_p} \geqslant \frac{E'^{\mathrm{PS}}_p - E_p}{E'_\gamma - E_p}, \tag{48}$$

and thus, by Eq. (47),

$$\frac{E'_p - E_p}{E'_\gamma - E_p} \geqslant \frac{E'^{\mathrm{PS}}_p - E_p}{E'_\gamma - E_p}. \tag{49}$$

In SA, $E_p \geqslant E_\gamma \geqslant E'_\gamma$. The above inequality gives us $E'^{\mathrm{PS}}_p \geqslant E'_p$. With the initial condition $E^{\mathrm{PS}}_p(0) = E_p(0)$, by induction, we proved that $\mu \geqslant \mu^{\mathrm{PS}}$ leads to $E^{\mathrm{PS}}_p \geqslant E_p$ all the time. ■

Knowing that $\mu \geqslant \mu^{\mathrm{PS}}$ does give us an upper bound on E_p, we further note that one can evaluate E^{PS}_p iteratively in Eq. (44) by using E_γ only. Replacing $E_p(k)$ with $E^{\mathrm{PS}}_p(k)$ in Eq. (42), we thus find the desired bound $\mathcal{I}(\kappa) \leqslant \mathcal{I}^{\mathrm{PS}}(\kappa)$.

Now the question is how to find such a μ^{PS} less than μ when μ is not accessible (since its calculation still requires E_p). Recall that μ gauges the speed of the relaxation dynamics. The relaxation time $\tau_{\mathrm{rel}}(P) \equiv 1/(1-\lambda_2(P))$ of the discrete Markov chain [31], defined by the second largest eigenvalue of the transition matrix P, $\lambda_2(P)$, should be consistent with μ. Therefore, we conjecture that the order of relaxation times of the two processes corresponding to μ and μ^{PS} is the inverse order of μ and μ^{PS}, i.e.,

$$\tau_{\mathrm{rel}} \leqslant \tau^{\mathrm{PS}}_{\mathrm{rel}} \Rightarrow \mu \geqslant \mu^{\mathrm{PS}}. \tag{50}$$

With this conjecture, we can compare $\lambda_2(P^{\mathrm{PS}})$ and $\lambda_2(P)$ of the transition matrices P^{PS} and P to find a condition on $\lambda_2(P^{\mathrm{PS}})$. The desired μ^{PS} can therefore be derived from $\lambda_2(P^{\mathrm{PS}})$ by the following proposition.

Proposition 2. For a partial swap model given by Eq. (43), $\lambda_2(P'^{\mathrm{PS}}) = 1 - \mu^{\mathrm{PS}}$.

Proof. Let N be the total number of states. The transition matrix P'^{PS} for such a PS model can be written explicitly as follows:

$$\begin{aligned} P'^{\mathrm{PS}} &= \begin{pmatrix} 1-\mu+\mu\gamma'_1 & \mu\gamma'_1 & \cdots & \mu\gamma'_1 \\ \mu\gamma'_2 & 1-\mu+\mu\gamma'_2 & \cdots & \mu\gamma'_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mu\gamma'_N & \mu\gamma'_N & \cdots & 1-\mu+\mu\gamma'_N \end{pmatrix} \\ &= (1-\mu)\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} + \mu \begin{pmatrix} \gamma'_1 & \cdots & \gamma'_1 \\ \gamma'_2 & \cdots & \gamma'_2 \\ \vdots & & \vdots \\ \gamma'_N & \cdots & \gamma'_N \end{pmatrix}, \end{aligned} \tag{51}$$

where we have neglected the superscript "PS" of the matrix elements. Note that the second matrix on the right-hand side has a rank of 1. By the rank-nullity theorem, for a matrix T, $\mathrm{Rank}(T) + \mathrm{Nullity}(T) = \dim(T)$, where $\mathrm{Nullity}(T) = \dim\{\vec{v}|T\vec{v} = 0\}$, the nullity of the second matrix is $N-1$, i.e., it has the eigenvalue of 0 with multiplicity of $N-1$ and the other eigenvalue is 1 given by its trace. Therefore, the eigenvalues of P'^{PS} are $\lambda_1(P'^{\mathrm{PS}}) = 1$ and $\lambda_2(P'^{\mathrm{PS}}) = \cdots = \lambda_N(P'^{\mathrm{PS}}) = 1 - \mu^{\mathrm{PS}}$. ■

For the transition matrix P in SA [Eq. (31)], the second largest eigenvalue $\lambda_2(P)$ can be bounded as [52]

$$\lambda_2(P) \leqslant 1 - \frac{\omega_{\min}}{\gamma_{\max}} \lambda_2(Q), \tag{52}$$

where $\omega_{\min} = \min_{i,j\in\mathcal{N}(i)} P_{ij}\gamma_j$ and $\gamma_{\max} = \max_i \gamma_i$. Here, Q is the Laplacian matrix associated with the state space [52],

$$Q_{ij} = \begin{cases} |\mathcal{N}(i)| & \text{if } i = j, \\ -1 & \text{if } j \in \mathcal{N}(i), \\ 0 & \text{otherwise}, \end{cases} \tag{53}$$

and $\lambda_2(Q)$ is the second smallest eigenvalue of Q. By Eq. (31),

$$\omega_{\min} = \min_{i,j\in\mathcal{N}(i)} \left(\frac{1}{|\mathcal{N}(j)|} \frac{\gamma_i\gamma_j}{\gamma_i + \gamma_j} \right) \geqslant \frac{1}{\mathcal{N}_{\max}} \frac{\gamma_{\min}}{2}, \tag{54}$$

where $\mathcal{N}_{\max} = \max_i |\mathcal{N}(i)|$ and $\gamma_{\min} = \min_i \gamma_i$. Substituting the above into Eq. (52), we have

$$\lambda_2(P) \leqslant 1 - \frac{\lambda_2(Q)}{2\mathcal{N}_{\max}} \exp(-\beta\Delta E_{\max}). \tag{55}$$

Since we want PS to be a slower process than SA, i.e., $\tau_{\mathrm{rel}}(P) \leqslant \tau_{\mathrm{rel}}(P^{\mathrm{PS}})$ followed by $\lambda_2(P) \leqslant \lambda_2(P^{\mathrm{PS}})$, we can simply set $\lambda_2(P'^{\mathrm{PS}}) = 1 - \mu^{\mathrm{PS}} = 1 - \frac{\lambda_2(Q)}{2\mathcal{N}_{\max}} \exp(-\beta'\Delta E_{\max})$, which gives us

$$\mu^{\mathrm{PS}}(k) = \frac{\lambda_2(Q)}{2\mathcal{N}_{\max}} \exp[-\beta(k+1)\Delta E_{\max}], \tag{56}$$

where we have recovered the k dependence for further uses. By the conjecture [Eq. (47)], such a $\mu^{\mathrm{PS}}(k)$ guarantees $\mu(k) \geqslant \mu^{\mathrm{PS}}(k)$. Hence, it leads to $\mathcal{I}(\kappa) \leqslant \mathcal{I}^{\mathrm{PS}}(\kappa)$ as we discussed.

For a concrete example of the $\mu^{\mathrm{PS}}(k)$ of Eq. (56), consider an n-spin Edwards-Anderson (EA) model [39], a generalization of the SK model with the Hamiltonian

$$H_{\mathrm{EA}} = \sum_{\langle k,l\rangle} g_{kl} s_k s_l, \tag{57}$$

where each spin $s_k \in \{-1,+1\}$ and $\langle k,l\rangle$ denotes the set of pairs of spins s_k and s_l having nonvanishing couplings g_{kl}. Here, we do not consider the self-energy terms, i.e. $g_{kk}=0$, $\forall k$. The system state i is a vector representing the spin configuration $\vec{s}^{(i)} = \{s_1^{(i)}, \dots, s_n^{(i)}\}$. The neighbors of i are configurations with only one spin orienting oppositely with respect to i. In other words, every neighborhood has the same size, i.e., $\mathcal{N}_{\max} = n$. Given that the total number of states is 2^n, the state space is an n-dimensional hypercube, whose $\lambda_2(Q) = 2$ [52]. In this case, Eq. (56) becomes

$$\mu^{\mathrm{PS}}(k) = \tfrac{1}{n}\exp[-\beta(k+1)\Delta E_{\max}]. \tag{58}$$

We will use this $\mu^{\mathrm{PS}}(k)$ to bound our numerical results in Sec. IV F.

D. Performance bound for simulated annealing

We now derive the performance bound for the simulated annealing we considered in this section.

We define the entropy production in the discrete-time Markov chain, denoted by $\Sigma(\kappa)$, as the entropy production in the corresponding continuous-time Markov chain. The detailed justification is given in the Appendix, with the key results listed below. The discrete-time analog of the speed limit in Eq. (25) is given by

$$\Sigma(\kappa) \geqslant \frac{(L_0 - L_\kappa)^2}{2\langle \mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa}. \tag{59}$$

Accordingly, Eq. (26) becomes

$$L_\kappa^2 \leqslant -\frac{(L_0 - L_\kappa)^2}{\langle \mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa} + 2\mathcal{I}(\kappa), \tag{60}$$

Substituting the conjectured bounds $\langle \mathcal{A}^{\mathrm{D}}\rangle_\kappa \leqslant 1/2$ and $\mathcal{I}(\kappa) \leqslant \mathcal{I}^{\mathrm{PS}}(\kappa)$ with the partial swap rate $\mu^{\mathrm{PS}}(k)$ in Eq. (56), we have

$$L_\kappa^2 \leqslant -\frac{2(L_0 - L_\kappa)^2}{\kappa} + 2\mathcal{I}^{\mathrm{PS}}(\kappa). \tag{61}$$

Solving this inequality for L_κ, we have the performance bound of SA,

$$L_\kappa \leqslant \frac{2L_0 + \sqrt{2\kappa\left[-L_0^2 + \mathcal{I}^{\mathrm{PS}}(\kappa)(\kappa+2)\right]}}{\kappa+2}. \tag{62}$$

Compared with the discrete-time analogy of the bound Eq. (27) (see the Appendix),

$$L_\kappa \leqslant \frac{L_0 + \sqrt{\langle \mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa\left(-L_0^2 + 2\mathcal{I}(\kappa)(\langle \mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa + 1)\right)}}{\langle \mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa + 1}, \tag{63}$$

Eq. (62) replaces the relative entropy term $\mathcal{I}$ with $\mathcal{I}^{\mathrm{PS}}$ associated with a partial swap model. Evaluating $\mathcal{I}^{\mathrm{PS}}(\kappa)$ does not require solving the evolution of $p(k)$ [Eq. (30)] but only the information on $E_\gamma(k)$, which can be obtained by knowing the energy spectrum of the system.

For specific models, the energy spectrum or the probability distribution of energy levels may be known, from which one can construct the partition function Z and find E_γ by $E_\gamma = -\partial \ln Z/\partial\beta$.

For systems with *a priori* unknown energy levels, one can sample the energy landscape to estimate the spectrum. The computational resources required in evaluating the bound on the error of the solution are also then in fact much lower than the resources required *for evaluating the error* L_κ via simulated annealing. To find L_κ, the distance between the statistical state $p(\kappa)$ and the equilibrium statistical state $\gamma(\kappa)$, directly through simulated annealing, many trajectories are indeed required, in order to obtain $p(k)$ at the kth step. A key advantage of Eq. (62) is that it replaces the need to evaluate that distance with the need to know the energy spectrum. Given a state space of size N, one can find the energy landscape by the order of N calls of the energy function. In contrast, for a faithful estimation of $p(k)$ via simulated annealing, the number of trajectories $N_{\mathrm{traj}} \gg N$ and each Monte Carlo time step for all trajectories needs N_{traj} calls of the energy function. Therefore, quantifying the computational cost by the number of calls of the energy function, evaluating our Eq. (62) is computationally much cheaper than performing simulated annealing to evaluate the error.

In the following, we exemplify the calculation of the bound for 1D Ising chains with known energy spectrum in Sec. IV E, and study the tightness of the bound by numerical simulation on SK models in Sec. IV F.

E. Example of a 1D Ising chain

In order to show how this bound can be analytically calculated, we consider the simulated annealing on a 1D n-spin Ising chain, which is a simplified EA model, with the Hamiltonian given by [53]

$$H_{\mathrm{1D}} = -J\sum_{i=1}^{n} s_i s_{i+1}, \tag{64}$$

where $s_i \in \{-1,+1\}$, J is the coupling constant and the periodic boundary condition is applied, i.e., $s_n s_{n+1} \equiv s_n s_1$. For even n, the partition function Z_{1D} at the inverse temperature β can be calculated as [53]

$$Z_{\mathrm{1D}} = 2^n[\cosh(\beta J)^n + \sinh(\beta J)^n]. \tag{65}$$

The equilibrium energy $E_{\gamma,\mathrm{1D}}$ is therefore

$$E_{\gamma,\mathrm{1D}} = -\frac{\partial \ln Z_{\mathrm{1D}}}{\partial\beta} = -\frac{Jn\cosh(\beta J)\sinh(\beta J)[\cosh(\beta J)^{n-2} + \sinh(\beta J)^{n-2}]}{\cosh(\beta J)^n + \sinh(\beta J)^n}. \tag{66}$$

Using the partial swap model [Eq. (44)], one can calculate $E^{\mathrm{PS}}_{p,1\mathrm{D}}(k) \equiv E^{\mathrm{PS}}_{p,1\mathrm{D}}(\beta(k))$ iteratively from $E_{\gamma,1\mathrm{D}}(k) \equiv E_{\gamma,1\mathrm{D}}(\beta(k))$ and find $\mathcal{I}^{\mathrm{PS}}_{1\mathrm{D}}(\kappa)$ by Eq. (42). On the other hand, given the energy landscape, the initial distance $L_0 = \sum_i |\gamma_i(0) - \gamma_i(\kappa)|$ is known. The bound Eq. (62) can thus be evaluated for this model.

Apart from evaluating Eq. (62) exactly, we can also deduce the scaling of L_κ for large κ. We consider the normal cooling schedules where the rate of varying temperature decreases with κ. Therefore, when κ gets large, $dk \equiv 1$ and $dE^{\mathrm{PS}}_{p,1\mathrm{D}}(k) \equiv E^{\mathrm{PS}}_{p,1\mathrm{D}}(k+1) - E^{\mathrm{PS}}_{p,1\mathrm{D}}(k)$ are relatively small. Equation (44) can be written as a differential equation:

$$\dot{E}^{\mathrm{PS}}_{p,1\mathrm{D}}(k) = -\mu^{\mathrm{PS}}_{1\mathrm{D}}(k)\big[E^{\mathrm{PS}}_{p,1\mathrm{D}}(k) - E_{\gamma,1\mathrm{D}}(k)\big], \tag{67}$$

where $\dot{x}(k) \equiv dx/dk$. This equation can be solved with the initial condition $E^{\mathrm{PS}}_{p,1\mathrm{D}}(0) = E_{\gamma,1\mathrm{D}}(0)$ by (see Supplemental Material of Ref. [21])

$$E^{\mathrm{PS}}_{p,1\mathrm{D}}(k) = E_{\gamma,1\mathrm{D}}(k) - \int_0^k e^{-\int_s^k \mu^{\mathrm{PS}}_{1\mathrm{D}}(s')ds'} \dot{E}_{\gamma,1\mathrm{D}}(s)ds. \tag{68}$$

Accordingly, for large κ, $\mathcal{I}^{\mathrm{PS}}_{1\mathrm{D}}(\kappa)$ calculated by Eq. (42) can be written as an integral in a form similar to Eq. (15),

$$\begin{aligned}\mathcal{I}^{\mathrm{PS}}_{1\mathrm{D}}(\kappa) &= \int_0^\kappa \big[E^{\mathrm{PS}}_{p,1\mathrm{D}}(k) - E_{\gamma,1\mathrm{D}}(k)\big]\dot{\beta}(k)dk \\ &= -\int_0^\kappa \int_0^k e^{-\int_s^k \mu^{\mathrm{PS}}_{1\mathrm{D}}(s')ds'} \dot{E}_{\gamma,1\mathrm{D}}(s)\dot{\beta}(k)ds\,dk,\end{aligned} \tag{69}$$

where the second line is from the substitution of Eq. (68). We note that

$$\dot{E}_\gamma = \sum_i E_i\dot{\gamma}_i = \sum_i E_i(E_\gamma - E_i)\gamma_i\dot{\beta} = -\sigma(E)^2_\gamma\,\dot{\beta}, \tag{70}$$

where the second equality used $\dot{\gamma}_i = (E_\gamma - E_i)\gamma_i\dot{\beta}$ [Eq. (13)], and $\sigma(E)^2_\gamma \equiv -\sum_i E_i(E_\gamma - E_i)\gamma_i$ is the energy variance for the thermal state γ. $\mathcal{I}^{\mathrm{PS}}_{1\mathrm{D}}(\kappa)$ can therefore be written as

$$\mathcal{I}^{\mathrm{PS}}_{1\mathrm{D}}(\kappa) = \int_0^\kappa \int_0^k \exp\left[-\int_s^k \mu^{\mathrm{PS}}_{1\mathrm{D}}(s')ds'\right]\sigma_{1\mathrm{D}}(E)^2_\gamma(s)\dot{\beta}(s)\dot{\beta}(k)ds\,dk, \tag{71}$$

where

$$\sigma_{1\mathrm{D}}(E)^2_\gamma = \frac{\partial^2 \ln Z_{1\mathrm{D}}}{\partial\beta^2} = \frac{4J^2n[\cosh(\beta J)^{2n}\sinh(\beta J)^2 - \cosh(\beta J)^2\sinh(\beta J)^{2n} + (n-1)(\cosh(\beta J)\sinh(\beta J))^n]}{[\cosh(\beta J)^n + \sinh(\beta J)^n]^2\sinh(2\beta J)^2}. \tag{72}$$

For large βJ, $\cosh(\beta J) \simeq \sinh(\beta J)$ and we have

$$\sigma_{1\mathrm{D}}(E)^2_\gamma \simeq \frac{J^2n(n-1)}{\sinh\,(2\beta J)^2} \simeq 4J^2n(n-1)e^{-4\beta J}. \tag{73}$$

Consequently, by Eq. (71),

$$\mathcal{I}^{\mathrm{PS}}_{1D}(\kappa) \simeq 4J^2n(n-1)\int_0^\kappa \int_0^k \exp\left[-\int_s^k \frac{1}{n}\exp[-2\beta(s')nJ]ds'\right]e^{-4\beta(s)J}\dot{\beta}(s)\dot{\beta}(k)ds\,dk. \tag{74}$$

where, as we suggested in Eq. (58), $\mu^{\mathrm{PS}}_{1\mathrm{D}}(k)$ is chosen as $\mu^{\mathrm{PS}}_{1\mathrm{D}}(k) \simeq \exp[-2\beta(k)nJ]/n$ with $\Delta E_{\max} = 2nJ$. Equation (74) hence provides a systematic way to calculate $\mathcal{I}^{\mathrm{PS}}_{1D}(\tau)$ for any annealing schedule $\beta(k)$.

For example, Consider the cooling schedule $\beta(k) = \beta_{\mathrm{i}} + (\beta_{\mathrm{f}} - \beta_{\mathrm{i}})k/\kappa$, and use the lowest partial swap rate $\bar{\mu} \equiv \exp[-2\beta_{\mathrm{f}}nJ]/n \leqslant \mu^{\mathrm{PS}}_{1\mathrm{D}}(k)$, $\forall k$, for simplicity. We then have

$$\mathcal{I}^{\mathrm{PS}}_{1D}(\kappa) \lesssim \frac{4J^2n(n-1)e^{-4\beta_{\mathrm{i}}J}(\beta_{\mathrm{f}} - \beta_{\mathrm{i}})^2}{-4J(\beta_{\mathrm{f}} - \beta_{\mathrm{i}}) + \bar{\mu}\kappa}\left[\frac{1 - e^{-4J(\beta_{\mathrm{f}} - \beta_{\mathrm{i}})}}{4J(\beta_{\mathrm{f}} - \beta_{\mathrm{i}})} + \frac{-1 + e^{-\bar{\mu}\kappa}}{\bar{\mu}\kappa}\right]. \tag{75}$$

Choosing $J = 1/n$ to fix the energy range $\Delta E_{\max} = 2$, we find $\mathcal{I}^{\mathrm{PS}}_{1D}(\kappa) \simeq \mathcal{O}(n\kappa^{-1})$. Substituting it into the bound Eq. (62), the scaling of the error L_κ is given as $L_\kappa \simeq \mathcal{O}(n^{1/2}\kappa^{-1/2})$. This result shows how the annealing performance for the 1D Ising chain is influenced by the system and protocol parameters, for this specific cooling schedule, which to our knowledge has not been derived before.

F. Numerical results for SK models

To verify the tightness of our bound in Eq. (62), we adopt an n-spin SK model to conduct SA. The Hamiltonian of an n-spin SK model is defined as [1,35]

$$H_{\mathrm{SK}} = \frac{1}{n^{3/2}}\sum_{k=1}^n\sum_{l=1}^n g_{kl}s_ks_l, \tag{76}$$

where $s_k \in \{-1, +1\}$ and $\{g_{kl}\}^n_{k,l=1}$ is a collection of independent and identically distributed standard Gaussian random variables. The normalization factor $1/n^{3/2}$ is chosen to ensure that the maximum energy has the order of magnitude of 1 [35]. The merit of this choice is that one can use the famous result of the SK model that $\lim_{n\to\infty}\mathbb{E}(H_{\mathrm{SK,max}}) \approx 0.76$, where $\mathbb{E}(\cdot)$ denotes the Gaussian expectation [54,55]. The partial swap rate in Eq. (58) is also applicable to the SK model. We thus

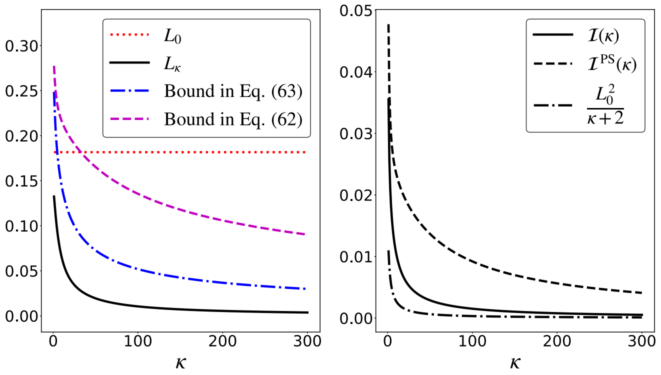

FIG. 3. Bounds on the performance of SA in solving a 7-spin SK model. $T_{\mathrm{i}} = 1.8$ is decreased inverse linearly to $T_{\mathrm{f}} = 0.8$ in κ steps. *Left:* The performance error L_κ and our bounds in Eqs. (63) and in (62), as well as the performance error of the initial state L_0. *Right:* The accumulated relative entropy due to temperature change $\mathcal{I}(\kappa)$ with our upper bound $\mathcal{I}^{\mathrm{PS}}(\kappa)$ and our lower bound from the tradeoff relation between $\mathcal{I}(\kappa)$ and $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa$ (for $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa \leqslant 1/2$).

have a nice estimation of $\Delta E_{\max}$ in the expression of $\mu^{\mathrm{PS}}(k)$ in the thermodynamic limit $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathbb{E}(\Delta E_{\max}) = \lim_{n\to\infty} 2\mathbb{E}(H_{\mathrm{SK},\max}) \approx 1.52. \tag{77}$$

This result with Eq. (58) provides a faithful way to find an appropriate $\mu^{\mathrm{PS}}(k)$ only in terms of cooling schedules for an SK model with large n.

In Fig. 3, we present the numerical results of SA conducted on a 7-spin SK model, in which the temperature is inverse-linearly decreased from $T_{\mathrm{i}} = 1.8$ to $T_{\mathrm{f}} = 0.8$ in κ steps. The left plot validates that both evolution history-dependent or -independent bounds [Eqs. (63) and (62)] provide upper limits for the performance error L_κ of SA. To reveal the κ dependence of the annealing-related quantity, the relative entropy from temperature descending $\mathcal{I}(\kappa)$, bounds from both directions are presented in the right plot, where $\mathcal{I}^{\mathrm{PS}}(\kappa)$ is calculated by the method suggested in Sec. IV C and the lower bound is given by the tradeoff relation in Eq. (28) with $\langle \mathcal{A}^{\mathrm{D}} \rangle_\kappa \leqslant 1/2$. Notably, the upper and lower bounds of $\mathcal{I}(\kappa)$ are saturated for fast (small κ) and slow (large κ) cooling, respectively, demonstrating the tightness of the two bounds.

For comparison, a similar simulation is provided in the accompanying paper [21], which adopts linear cooling schedules.

G. Relaxation rate μ of spin glasses in SA

From the previous discussion, we have seen that the choice of $\mu^{\mathrm{PS}}(k)$ in Eq. (58) does give us an evolution history-independent bound for the example of an SK model. However, as is shown in Fig. 3, there is still a gap between the history-independent bound in Eq. (62) and the history-dependent bound in Eq. (63). A natural question to ask is whether Eq. (58) is an optimal choice of $\mu^{\mathrm{PS}}(k)$ that satisfies $\mu(k) \geqslant \mu^{\mathrm{PS}}(k)$. Can we find the largest $\mu^{\mathrm{PS}}(k)$ allowed? In this section, we will pursue this possibility further by looking at the relaxation rate $\mu(k)$ defined in Eq. (46) in more detail.

Referring to the time evolution of SA in Eq. (30), we have (again omitting the arguments k and using $'$ to label quantities at $k+1$ step)

$$p_i' = p_i + \frac{1}{n} \sum_{j\in\mathcal{N}(i)} (A_{ij}' p_j - A_{ji}' p_i), \tag{78}$$

where we have considered the equal-sized neighborhoods, $|\mathcal{N}(i)| = n$, $\forall i$, and the acceptance rate A_{ij} is given by Eq. (32). Thus, E_p' is given by

$$\begin{aligned} E_p' &= \sum_i E_i p_i' \\ &= E_p + \frac{1}{n} \sum_i \sum_{j\in\mathcal{N}(i)} (E_i A_{ij}' p_j - E_i A_{ji}' p_i) \\ &= E_p + \frac{1}{n} \sum_i \sum_{j\in\mathcal{N}(i)} (E_j - E_i) A_{ji}' p_i \\ &= E_p + \frac{1}{n} \sum_i \sum_{j\in\mathcal{N}(i)} \Delta E_{ji} \frac{\gamma_j' p_i}{\gamma_i' + \gamma_j'}, \end{aligned} \tag{79}$$

where we have defined $\Delta E_{ji} \equiv E_j - E_i$. Then

$$E_p' - E_p = \frac{1}{n} \sum_i \sum_{j\in\mathcal{N}(i)} \Delta E_{ji} \frac{\gamma_j'}{\gamma_i' + \gamma_j'} (p_i - \gamma_i'). \tag{80}$$

To relate it with μ, we want to rewrite the right-hand side of Eq. (80) in terms of $E_\gamma' - E_p$. The following proposition is thus considered.

Proposition 3. For an n-spin EA model with the Hamiltonian defined in Eq. (57), the following holds:

$$\sum_{j\in\mathcal{N}(i)} \Delta E_{ji} = -4E_i. \tag{81}$$

Proof. We rewrite the Hamiltonian in Eq. (57) as

$$\begin{aligned} H_{\mathrm{EA}} &= \sum_{\langle k,l \rangle} g_{kl} s_k s_l \\ &= \frac{1}{2} \sum_{k=1}^{n} \sum_{l\in\mathcal{C}(k)} g_{kl} s_k s_l \\ &\equiv \frac{1}{2} \sum_{k=1}^{n} s_k h_k, \end{aligned} \tag{82}$$

where $\mathcal{C}(k)$ is the set of spins interacting with spin s_k and $h_k \equiv \sum_{l\in\mathcal{C}(k)} g_{kl} s_l$ is the effective field applied on s_k. The factor $1/2$ is from double counting. Now consider a spin configuration (state i) $\vec{s}^{(i)} = \{s_1^{(i)}, \ldots, s_n^{(i)}\}$ with energy $E_i = \sum_{k=1}^{n} s_k^{(i)} h_k^{(i)}/2$. A neighboring state j can be generated by flipping the jth spin in $\vec{s}^{(i)}$, i.e., $\vec{s}^{(j)} = \{s_1^{(i)}, \ldots, -s_j^{(i)}, \ldots, s_n^{(i)}\}$. Therefore, the energy difference ΔE_{ji} is

$$\begin{aligned} \Delta E_{ji} &= E_j - E_i \\ &= E_{\mathrm{rest},j} - s_j^{(i)} h_j^{(i)} - \left(E_{\mathrm{rest},j} + s_j^{(i)} h_j^{(i)}\right) \\ &= -2 s_j^{(i)} h_j^{(i)}, \end{aligned} \tag{83}$$

where $E_{\text{rest},j}$ is the energy not involving the interactions with the spin s_j. Summing over all neighbors of i, we find that

$$\sum_{j\in\mathcal{N}(i)} \Delta E_{ji} = \sum_{j=1}^{n} -2 s_j^{(i)} h_j^{(i)} = -4E_i. \tag{84}$$

■

As a result of this proposition, we have

$$E'_\gamma - E_p = \frac{1}{4}\sum_i \sum_{j\in\mathcal{N}(i)} \Delta E_{ji}(p_i - \gamma'_i). \tag{85}$$

Therefore, Eq. (80) can be written as

$$\begin{aligned} E'_p - E_p &= \frac{1}{n}\sum_i \sum_{j\in\mathcal{N}(i)} \frac{\Delta E_{ji}}{2}(p_i - \gamma'_i) + \frac{1}{n}\sum_i \sum_{j\in\mathcal{N}(i)} \Delta E_{ji}\left(\frac{\gamma'_j}{\gamma'_i + \gamma'_j} - \frac{1}{2}\right)(p_i - \gamma'_i) \\ &= \frac{2}{n}(E'_\gamma - E_p) - \frac{1}{n}\sum_i \sum_{j\in\mathcal{N}(i)} \frac{\Delta E_{ji}}{2} \tanh\left(\beta' \frac{\Delta E_{ji}}{2}\right)(p_i - \gamma'_i). \end{aligned} \tag{86}$$

We then obtain

$$\begin{aligned} \mu &:= \frac{E'_p - E_p}{E'_\gamma - E_p} \\ &= \frac{2}{n} - \frac{1}{n}\sum_i \sum_{j\in\mathcal{N}(i)} \frac{\Delta E_{ji}}{2} \tanh\left(\beta' \frac{\Delta E_{ji}}{2}\right) \frac{p_i - \gamma'_i}{E'_\gamma - E_p}. \end{aligned} \tag{87}$$

At high temperatures where $\beta'\Delta E_{\max} \ll 1$ with $\Delta E_{\max} = \max_i E_i - \min_i E_i$, $\mu \simeq 2/n$. This reflects the fact that, in the high-temperature limit, the complex structure of the energy landscape can be neglected and the relaxation rate is only diminished by the size of the system. As temperature decreases, the second term on the right-hand side of Eq. (87) becomes significant and therefore introduces more complication to the evaluation of μ. We note that according to the cooling schedule, $\beta(k)$, $\mu(k(\beta)) \equiv \mu(\beta)$ can be viewed as a function of β. If $\mu(k)$ does not depend on the system state $p(k)$, $\mu(\beta)$ should be the same for all cooling schedules when β is fixed. However, Fig. 4 provides a counterexample of a 7-spin SK model showing that $\mu(k)$ is not a sole function of β but relies on the history of β. Hence, finding $\mu(k)$ without referring to $p(k)$ seems intractable. Moreover, as is depicted in the red dash-dotted line in Fig. 4, the μ^{PS} in Eq. (58) does bound μ irrespective of cooling schedules and is even saturated for large β. Therefore, at least for this specific example, at low temperatures, we cannot do much better than the μ^{PS} in Eq. (58). On the other hand, Fig. 4 provides numerical evidence for our conjecture [Eq. (50)].

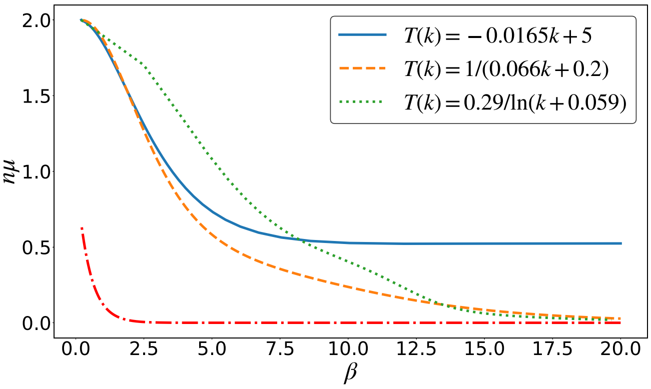

FIG. 4. The relaxation rate $\mu(\beta)$ (multiplied with n) in the SA of a 7-spin SK model. The simulation takes 300 steps with temperature changing from $T_{\text{i}} = 5$ to $T_{\text{f}} = 0.05$, according to the linear (blue solid line), inversely linear (orange dashed line) and inversely logarithmic (green dotted line) cooling schedules, respectively. The partial swap rate $\mu^{\text{PS}}(\beta)$ in Eq. (62) multiplied with n is represented by the red dash-dotted line.

Finally, we note that the asymptotic behavior of $\mu \to 2/n$ at high temperatures, $\beta \to 0$, is recognized in Fig. 4 for all three cooling schedules. To take advantage of this property, one may use $\mu^{\text{PS}} = 2/n$ for high temperatures and then use the μ^{PS} in Eq. (58) in further annealing to obtain a tighter bound than Eq. (62).

V. CONCLUSION

In this article, we derived a general bound for quantifying the annealing performance. The bound is derived in the framework of finite-time stochastic thermodynamics and thus can be applied to nanoscale systems and particular quantum systems whose evolution satisfies stochastic master equations. The bound shows that the 1-norm distance between the final statistical state of the system and the final thermal state is restricted by two problem-specific parameters: the averaged activity $\langle\mathcal{A}\rangle_\tau$ and the change of relative entropy due to the annealing $\mathcal{I}(\tau)$. The two quantities prove to obey a tradeoff relation forced by the initial distance L_0.

We further applied this bound to study the finite-time performance of simulated annealing. Under a reasonable conjecture that the discrete activity $\mathcal{A}^{\text{D}} \leqslant 1/2$ and purposing a slower partial swap model to bound $\mathcal{I} \leqslant \mathcal{I}^{\text{PS}}$, we obtained a modified bound which does not require the intermediate states of the system and only refers to the rule of temperature descent and the energy landscape. Consequently, the new bound applies to any Glauber dynamics SA schedules and any cost function. We employed 1D Ising chains as an example to illustrate how the bound can be evaluated analytically. To validate our conjectures, we presented numerical simulations of 7-spin SK models, demonstrating the performance of our bounds. Above all, our results show that recent developments in stochastic thermodynamics can be adapted to characterize the finite time behavior of SA, and we expect such approaches can be further developed to understand other Monte Carlo algorithms.

ACKNOWLEDGMENT

We gratefully acknowledge valuable discussions with Alexander Yosifov, Barry Sanders, Li Xiao, and Yu Chai. This work was supported by National Natural Science Foundation of China (Grants No. 12050410246 and No. 12005091).

APPENDIX: FROM CONTINUOUS-TIME ANNEALING TO DISCRETE-TIME SA

Here we prove that the discrete-time evolution in SA, i.e., Eq. (30) in the main text, can always be simulated by a continuous-time master equation defined in Eq. (2) in the main text. For the clarity of notations, below we write probabilities in the discrete-time with a superscript D and probabilities in the continuous time with a superscript C. Then, our aim is to prove that

$$p_i^{\mathrm{D}}(k+1)=\sum_j P_{ij}(k+1)p_j^{\mathrm{D}}(k), \tag{A1}$$

can be constructed by

$$\dot{p}_i^{\mathrm{C}}(t)=\sum_j \Gamma_{ij}(t)p_j^{\mathrm{C}}(t), \tag{A2}$$

such that

$$p_i^{\mathrm{D}}(k)=p_i^{\mathrm{C}}(t_k) \tag{A3}$$

for times $t_0=0,t_1,\dots,t_k,\dots,t_\kappa=\tau$ with κ denoting the number of iterations in SA.

Integrating Eq. (A2) from time t_k to time t_{k+1}, we obtain

$$\begin{aligned} p_i^{\mathrm{C}}(t_{k+1}) &= p_i^{\mathrm{C}}(t_k)+\int_{t_k}^{t_{k+1}} dt\sum_j \Gamma_{ij}(t)p_j^{\mathrm{C}}(t) \\ &= \sum_j W_{ij}(k+1)p_j^{\mathrm{C}}(t_k), \end{aligned} \tag{A4}$$

where we have defined

$$W_{ij}(k+1)=\delta_{ij}+\int_{t_k}^{t_{k+1}} dt\, \Gamma_{ij}(t)p_j^{\mathrm{C}}(t)/p_j^{\mathrm{C}}(t_k) \tag{A5}$$

for each $k=0,1,\dots,\kappa-1$, where δ_{ij} is the Kronecker delta. To show that there always exists a valid transition rate matrix making Eq. (A4) reproduce Eq. (A1), we introduce

$$\Gamma_{ij}(t)=\sum_{k=0}^{\kappa-1}\delta(t-t_k^+)P_{ij}(k+1) \tag{A6}$$

for $i\neq j$ and $\Gamma_{ii}(t)=-\sum_{j(\neq i)}\Gamma_{ji}(t)$. Here, $\delta(\cdot)$ is the Dirac δ function and $t_k^+>t_k$ is a value close to t_k.

As we will show, $P_{ij}(k+1)$ guarantees that the transition rate defined in Eq. (A6) induces the required thermal states and detailed balance relation. Due to the relation of $P_{ij}(k+1)$ demanded by SA,

$$P_{ij}(k+1)\gamma_j^{\mathrm{D}}(k+1)=P_{ji}(k+1)\gamma_i^{\mathrm{D}}(k+1), \tag{A7}$$

it can be verified that the state defined as

$$\gamma^{\mathrm{C}}(t)=\gamma^{\mathrm{D}}(k+1) \quad \text{if } t_k^+<t\leqslant t_{k+1} \tag{A8}$$

satisfies

$$\Gamma_{ij}(t)\gamma_j^{\mathrm{C}}(t)=\Gamma_{ji}(t)\gamma_i^{\mathrm{C}}(t) \tag{A9}$$

for all $i\neq j$. Thus, $\gamma^{\mathrm{C}}(t)$ always satisfies the detailed balance condition and is a valid equilibrium state for the process governed by Eq. (A6). Meanwhile, substituting Eq. (A6) into Eq. (A5), we obtain

$$\begin{aligned} W_{ij}(k+1) &= P_{ij}(k+1),\ \forall i\neq j, \\ W_{ii}(k+1) &= P_{ii}(k+1),\ \forall i. \end{aligned} \tag{A10}$$

where we have used the fact that $p_j^{\mathrm{C}}(t_k^+)/p_j^{\mathrm{C}}(t_k)\to 1$.

Finally, by taking $p_i^{\mathrm{D}}(0)=p_i^{\mathrm{C}}(t_0)$, we obtain that the continuous-time evolution described by Eqs. (A2) and (A6) is in accordance with the discrete-time evolution described by Eqs. (A1) and (A7) at discrete times $t_0,\dots,t_\kappa$.

Consequently, the time-averaged activity $\langle\mathcal{A}^{\mathrm{C}}\rangle_\tau$ is defined as

$$\begin{aligned} \langle\mathcal{A}^{\mathrm{C}}\rangle_\tau &= \frac{1}{\tau}\int_0^\tau dt\sum_i\sum_{j(\neq i)}\Gamma_{ij}(t)p_j^{\mathrm{C}}(t) \\ &= \frac{1}{\tau}\sum_{k=0}^{\kappa-1}\int_{t_k}^{t_{k+1}} dt\sum_i\sum_{j(\neq i)}\Gamma_{ij}(t)p_j^{\mathrm{C}}(t) \\ &= \frac{1}{\tau}\sum_{k=0}^{\kappa-1}\sum_i\sum_{j(\neq i)}P_{ij}(k+1)p_j^{\mathrm{D}}(k). \end{aligned} \tag{A11}$$

To obtain an analogy of $\langle\mathcal{A}^{\mathrm{C}}\rangle_\tau$ for the discrete case in SA, we suppose that the time interval $t_{k+1}-t_k$ is a unit time 1, such that $\tau=\kappa$, and define the *iteration-averaged* activity $\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa$ as

$$\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa = \langle\mathcal{A}^{\mathrm{C}}\rangle_\tau \tag{A12}$$

$$= \frac{1}{\kappa}\sum_{k=0}^{\kappa-1}\sum_i\sum_{j(\neq i)}P_{ij}(k+1)p_j^{\mathrm{D}}(k) \tag{A13}$$

$$\equiv \frac{1}{\kappa}\sum_{k=0}^{\kappa-1}\mathcal{A}^{\mathrm{D}}(k), \tag{A14}$$

where the discrete activity is defined as

$$\mathcal{A}^{\mathrm{D}}(k)=\sum_i\sum_{j(\neq i)}P_{ij}(k+1)p_j^{\mathrm{D}}(k). \tag{A15}$$

Since the entropy production depends on the path of the statistical state of the system, it is more sensible to define the entropy production for the discrete-time evolution according to the underlying continuous-time evolution, in which the entropy production rate is well defined as Eq. (18) in the main text, i.e.,

$$\dot{\Sigma}^{\mathrm{C}}=\frac{1}{2}\sum_i\sum_{j(\neq i)}\left[\Gamma_{ij}p_j^{\mathrm{C}}-\Gamma_{ji}p_i^{\mathrm{C}}\right]\ln\frac{\Gamma_{ij}p_j^{\mathrm{C}}}{\Gamma_{ji}p_i^{\mathrm{C}}}. \tag{A16}$$

Therefore, we define

$$\Sigma^{\mathrm{D}}(\kappa)\equiv\Sigma^{\mathrm{C}}(\tau)=\int_0^\tau \dot{\Sigma}^{\mathrm{C}}(t)dt. \tag{A17}$$

By the "speed limit" with $\Sigma^{\mathrm{C}}(\tau)$ in Eq. (25) in the main text, we have

$$\Sigma^{\mathrm{D}}(\kappa)=\Sigma^{\mathrm{C}}(\tau)\geqslant\frac{(L_0-L_\tau)^2}{2\langle\mathcal{A}^{\mathrm{C}}\rangle_\tau\tau}=\frac{(L_0-L_\kappa)^2}{2\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa\kappa}, \tag{A18}$$

where we have used $\tau = \kappa$, Eq. (A14) and

$$L_0 = \sum_i \left|p_i^{\mathrm{C}}(0) - \gamma_i^{\mathrm{C}}(\tau)\right| = \sum_i \left|p_i^{\mathrm{D}}(0) - \gamma_i^{\mathrm{D}}(\kappa)\right|, \tag{A19}$$

$$L_\tau = \sum_i \left|p_i^{\mathrm{C}}(\tau) - \gamma_i^{\mathrm{C}}(\tau)\right|, \tag{A20}$$

$$L_\kappa = \sum_i \left|p_i^{\mathrm{D}}(\kappa) - \gamma_i^{\mathrm{D}}(\kappa)\right| = L_\tau. \tag{A21}$$

The last equality $L_\kappa = L_\tau$ is from $p^{\mathrm{D}}(\kappa) = p^{\mathrm{C}}(\tau)$ by Eq. (A3) and $\gamma^{\mathrm{D}}(\kappa) = \gamma^{\mathrm{C}}(\tau)$ by Eq. (A8). Eq. (A18) is therefore a “speed limit’ in the discrete-time case.

Accordingly, we can define the accumulated change in relative entropy $\mathcal{I}^{\mathrm{D}}(\kappa)$ by $\mathcal{I}^{\mathrm{C}}(\tau)$ in Eq. (15) in the main text, such that

$$\begin{aligned}
\mathcal{I}^{\mathrm{D}}(\kappa) &\equiv \mathcal{I}^{\mathrm{C}}(\tau) && \text{(A22)}\\
&= \int_0^\tau \left[E_p^{\mathrm{C}}(t) - E_\gamma^{\mathrm{C}}(t)\right]\dot{\beta}(t)dt && \text{(A23)}\\
&= \sum_i \int_0^\tau E_i\left[p_i^{\mathrm{C}}(t) - \gamma_i^{\mathrm{C}}(t)\right]\dot{\beta}(t)dt && \text{(A24)}\\
&= \sum_{k=0}^{\kappa-1} \sum_i \int_{t_k}^{t_{k+1}} E_i\left[p_i^{\mathrm{C}}(t) - \gamma_i^{\mathrm{C}}(t)\right]\dot{\beta}(t)dt && \text{(A25)}\\
&= \sum_{k=0}^{\kappa-1} \sum_i E_i\left[p_i^{\mathrm{C}}(t_{k+1}) - \gamma_i^{\mathrm{C}}(t_{k+1})\right]\\
&\quad \times [\beta(t_{k+1}) - \beta(t_k)] && \text{(A26)}\\
&= \sum_{k=0}^{\kappa-1} \sum_i E_i\left[p_i^{\mathrm{D}}(k+1) - \gamma_i^{\mathrm{D}}(k+1)\right]\\
&\quad \times [\beta(k+1) - \beta(k)] && \text{(A27)}\\
&= \sum_{k=0}^{\kappa-1} \left[E_p^{\mathrm{D}}(k+1) - E_\gamma^{\mathrm{D}}(k+1)\right]\\
&\quad \times [\beta(k+1) - \beta(k)], && \text{(A28)}
\end{aligned}$$

where in the fourth line we used the fact that $p^{\mathrm{C}}(t) = p^{\mathrm{C}}(t_{k+1})$ and $\gamma^{\mathrm{C}}(t) = \gamma^{\mathrm{C}}(t_{k+1})$ for $t_k^+ < t \leqslant t_{k+1}$, while in the fifth line $\beta(t_k) = \beta(k)$ is in accordance with Eq. (A8).

From Eq. (16) in the main text, we have

$$\begin{aligned}
S(p^{\mathrm{D}}(\kappa)||\gamma^{\mathrm{D}}(\kappa)) &= S(p^{\mathrm{C}}(\tau)||\gamma^{\mathrm{C}}(\tau)) && \text{(A29)}\\
&= -\Sigma^{\mathrm{C}}(\tau) + \mathcal{I}^{\mathrm{C}}(\tau) && \text{(A30)}\\
&= -\Sigma^{\mathrm{D}}(\kappa) + \mathcal{I}^{\mathrm{D}}(\kappa). && \text{(A31)}
\end{aligned}$$

Using Pinsker’s inequality, $L_\kappa^2 \leqslant 2S(p^{\mathrm{D}}(\kappa)||\gamma^{\mathrm{D}}(\kappa))$, and the “speed limit” [Eq. (A18)], we obtain

$$L_\kappa^2 \leqslant -\frac{(L_0 - L_\kappa)^2}{\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa} + 2\mathcal{I}^{\mathrm{D}}(\kappa), \tag{A32}$$

which is an analogy of Eq. (26) in the main text. Solving the inequality for L_κ gives us the discrete-time performance bound:

$$L_\kappa \leqslant \frac{L_0 + \sqrt{\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa\left(-L_0^2 + 2\mathcal{I}^{\mathrm{D}}(\kappa)(\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa + 1)\right)}}{\langle\mathcal{A}^{\mathrm{D}}\rangle_\kappa \kappa + 1}, \tag{A33}$$

analogous to Eq. (27) in the main text.

[1] S. Kirkpatrick, Jr., C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, Science **220**, 671 (1983).

[2] D. Bertsimas and J. Tsitsiklis, Simulated annealing, Stat. Sci. **8**, 10 (1993).

[3] L. Ingber, Simulated annealing: Practice versus theory, Math. Comput. Modell. **18**, 29 (1993).

[4] P. Salamon, P. Sibani, and R. Frost, *Facts, Conjectures, and Improvements for Simulated Annealing* (SIAM, Philadelphia, 2002).

[5] W. D. Callister and D. G. Rethwisch, *Materials Science and Engineering: An Introduction*, 10th ed. (Wiley, New York, 2018).

[6] R. C. Bernardi, M. C. Melo, and K. Schulten, Enhanced sampling techniques in molecular dynamics simulations of biological systems, Biochim. Biophys. Acta, Gen. Subj. **1850**, 872 (2015), special issue, Recent Developments of Molecular Dynamics.

[7] O. Ekren and B. Y. Ekren, Size optimization of a PV/wind hybrid energy conversion system with battery storage using simulated annealing, Appl. Energy **87**, 592 (2010).

[8] R. Meiri and J. Zahavi, Using simulated annealing to optimize the feature selection problem in marketing applications, Eur. J. Oper. Res. **171**, 842 (2006).

[9] F. Cai, S. Kumar, T. V. Vaerenbergh, X. Sheng, R. Liu, C. Li, Z. Liu, M. Foltin, S. Yu, Q. Xia, J. J. Yang, R. Beausoleil, W. D. Lu, and J. P. Strachan, Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks, Nat. Electron. **3**, 409 (2020).

[10] N. Mohseni, P. L. McMahon, and T. Byrnes, Ising machines as hardware solvers of combinatorial optimization problems, Nat. Rev. Phys. **4**, 363 (2022).

[11] B. Hajek, Cooling schedules for optimal annealing, Math. Oper. Res. **13**, 311 (1988).

[12] F. Romeo and A. Sangiovanni-Vincentelli, A theoretical framework for simulated annealing, Algorithmica **6**, 302 (1991).

[13] B. Andresen, Finite-time thermodynamics and simulated annealing, in *Entropy and Entropy Generation: Fundamentals and Applications*, edited by J. S. Shiner (Springer Netherlands, Dordrecht, 1996), pp. 111–127.

[14] M. P. Desai, Some results characterizing the finite time behaviour of the simulated annealing algorithm, Sadhana **24**, 317 (1999).

[15] P. N. Strenski and S. Kirkpatrick, Analysis of finite length annealing schedules, Algorithmica **6**, 346 (1991).

[16] Y. Nourani and B. Andresen, A comparison of simulated annealing cooling strategies, J. Phys. A: Math. Gen. **31**, 8373 (1998).

[17] A. Nolte and R. Schrader, A note on the finite time behavior of simulated annealing, Math. Oper. Res. **25**, 476 (2000).

[18] E. Fontenas and O. François, Improved bounds for the large-time behaviour of simulated annealing, J. Appl. Probab. **40**, 961 (2003).

[19] A. A. Albrecht, A problem-specific convergence bound for simulated annealing-based local search, Lect. Notes Comput. Sci. **3045**, 405 (2004).

[20] A. Nolte and R. Schrader, Simulated annealing and its problems to color graphs, in *Algorithms — ESA'96*, edited by J. Diaz and M. Serna (Springer, Berlin, 1996), pp. 138–151.

[21] Y. Luo, Y.-Z. Zhen, X. Liu, D. Ebler, and O. Dahlsten, preceding paper, General limit to thermodynamic annealing performance, Phys. Rev. E **108**, L052105 (2023).

[22] N. Shiraishi, K. Funo, and K. Saito, Speed limit for classical stochastic processes, Phys. Rev. Lett. **121**, 070601 (2018).

[23] A. Dechant, Minimum entropy production, detailed balance and Wasserstein distance for continuous-time Markov processes, J. Phys. A: Math. Theor. **55**, 094001 (2021).

[24] T. Van Vu and K. Saito, Thermodynamic unification of optimal transport: Thermodynamic uncertainty relation, minimum dissipation, and thermodynamic speed limits, Phys. Rev. X **13**, 011013 (2023).

[25] J. M. R. Parrondo, C. V. den Broeck, and R. Kawai, Entropy production and the arrow of time, New J. Phys. **11**, 073008 (2009).

[26] M. Esposito and C. V. den Broeck, Three faces of the second law: I. Master equation formulation, Phys. Rev. E **82**, 011143 (2010).

[27] M. Esposito and C. V. D. Broeck, Second law and Landauer principle far from equilibrium, Europhys. Lett. **95**, 40004 (2011).

[28] Y.-Z. Zhen, D. Egloff, K. Modi, and O. Dahlsten, Universal bound on energy cost of bit reset in finite time, Phys. Rev. Lett. **127**, 190602 (2021).

[29] T. Schmiedl and U. Seifert, Optimal finite-time processes in stochastic thermodynamics, Phys. Rev. Lett. **98**, 108301 (2007).

[30] V. N. Kampen, Stochastic *Processes in Physics and Chemistry* (Elsevier, Amsterdam, 2007).

[31] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, 2nd ed. (American Mathematical Society, Providence, 2017).

[32] I. Prigogine and P. Van Rysselberghe, Introduction to thermodynamics of irreversible processes, J. Electrochem. Soc. **110**, 97C (1963).

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (John Wiley & Sons, Hoboken, New Jersey, 2006).

[34] M. Pinsker and A. Feinstein, *Information and Information Stability of Random Variables and Processes*, Holden-Day series in time series analysis (Holden-Day, 1964).

[35] D. Panchenko, *The Sherrington-Kirkpatrick Model* (Springer, New York, 2013).

[36] R. J. Glauber, Time-dependent statistics of the Ising model, J. Math. Phys. **4**, 294 (1963).

[37] D. A. Levin, Malwina, J. Luczak, Y. Peres, D. A. Levin, M. J. Luczak, and Y. Peres, Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability, Probab. Theory Relat. Fields **146**, 223 (2010).

[38] J. C. Walter and G. T. Barkema, An introduction to Monte Carlo methods, Physica A **418**, 78 (2015).

[39] S. F. Edwards and P. W. Anderson, Theory of spin glasses, J. Phys. F **5**, 965 (1975).

[40] W. Pusz and S. L. Woronowicz, Passive states and KMS states for general quantum systems, Commun. Math. Phys. **58**, 273 (1978).

[41] A. Lenard, Thermodynamical proof of the Gibbs formula for elementary quantum systems, J. Stat. Phys. **19**, 575 (1978).

[42] M. Perarnau-Llobet, K. V. Hovhannisyan, M. Huber, P. Skrzypczyk, J. Tura, and A. Acin, Most energetic passive states, Phys. Rev. E **92**, 042147 (2015).

[43] C. Sparaciari, D. Jennings, and J. Oppenheim, Energetic instability of passive states in thermodynamics, Nat. Commun. **8**, 1895 (2017).

[44] N. Koukoulekidis, R. Alexander, T. Hebdige, and D. Jennings, The geometry of passivity for quantum systems and a novel elementary derivation of the Gibbs state, Quantum **5**, 411 (2021).

[45] V. Scarani, M. Ziman, P. Stelmachovic, N. Gisin, and V. Buzek, Thermalizing quantum machines: Dissipation and entanglement, Phys. Rev. Lett. **88**, 097905 (2002).

[46] C. Browne, A. J. P. Garner, O. C. O. Dahlsten, and V. Vedral, Guaranteed energy-efficient bit reset in finite time, Phys. Rev. Lett. **113**, 100603 (2014).

[47] J. D. Nulton and P. Salamon, Statistical mechanics of combinatorial optimization, Phys. Rev. A **37**, 1351 (1988).

[48] P. Salamon, J. D. Nulton, J. R. Harland, J. Pedersen, G. Ruppeiner, and L. Liao, Simulated annealing with constant thermodynamic speed, Comput. Phys. Commun. **49**, 423 (1988).

[49] B. Andresen and J. M. Gordon, Constant thermodynamic speed for minimizing entropy production in thermodynamic processes and simulated annealing, Phys. Rev. E **50**, 4346 (1994).

[50] G. Ruppeiner, J. M. Pedersen, and P. Salamon, Ensemble approach to simulated annealing, J. Phys. (Paris) **1**, 455 (1991).

[51] B. Anderson, Finite-time thermodynamics and simulated annealing, in *Entropy and Entropy Generation: Fundamentals and Applications* (Springer, Berlin, 2002), p. 111.

[52] M. P. Desai and V. B. Rao, On the convergence of reversible Markov chains, SIAM J. Matrix Anal. Appl. **14**, 950 (1993).

[53] E. Burovski, W. Janke, M. Guskova, and L. Shchur, Acceptance rate is a thermodynamic function in local Monte Carlo algorithms, Phys. Rev. E **100**, 063303 (2019).

[54] G. Parisi, A sequence of approximated solutions to the S-K model for spin glasses, J. Phys. A: Math. Gen. **13**, L115 (1980).

[55] D. Panchenko, Introduction to the SK model, arXiv:1412.0170.